Analysing the Danish public opinion about vegetarianism in Danish newspapers using sentiment analysis

Sebastian Scott Engen (snuggle.hogwarts@gmail.com)
School of Communication and Cognition, University of Aarhus,
Snogebæksvej 41, 8210 Aarhus, Denmark

Nina Mathilde Dyrberg (ninadyrberg@hotmail.com)
School of Communication and Cognition, University of Aarhus,
Ceresbyen 10 C, 2, 4, 8000 Aarhus C, Denmark

Abstract

Traditionally meat has been on the table in Denmark, however a rise in vegetarian products and a decline in meat consumption is changing the landscape of the breakfast table in the 21st century.(Danskernes madvaner, 2006; Økonomisk analyse, 2014)
Hereof this study seeks to investigate the development of Danish public opinion concerning vegetarianism over the last 10 years using sentiment analysis. 400 articles are gathered from the two newspapers 'Information' and 'Jyllandsposten' respectively using web-mining with an equal number of articles concerning carnism[1] and vegetarianism[2]. A sentiment analysis is accomplished using the Danish dictionary AFINN to score words in the range of a positive (5) to a negative (-5) (AFINN, 2018). Subsequently the corpus is analyzed using linear regression to create two different interaction models with different predictors, yet no significant interaction effects are found for either model. Additionally, this paper examines whether there has been an increase in coverage across all national Danish newspapers over the last 10 years, but once again no significant linear development is found. Keeping the nonsignificant results in mind this paper discusses its several methodological limitations, along with suggestions for future research.

Keywords

Sentiment analysis, News analysis, AFINN, Opinion Mining, Text Mining

1. Introduction[1]

In a late modern society flooded with information from all its new gizmos and gadgets, the global society is getting increasingly focused on the currency of attention. Considering this, it is interesting to investigate how the media landscape is shaping, and how it renders and reflects the public opinion through what articles it writes and how it writes them.

An ideal newspaper is expected to be objective (Journalistic Objectivity, 2018), and although absolute objectivity is practically impossible, a study on the two danish newspapers 'Information' and 'Jyllandsposten' showed remarkable difference in the coverage of the danish invasion of Iraq in 2003 (Hjarvard, 2007); the right-winged newspaper 'Jyllandsposten' leading a very supportive narrative of the invasion while the left-winged newspaper 'Information', in contrast, remained highly critical in its framing[3].

Motivated by this and the fact that national newspapers are seeing a steady decline in their readerships (Færre avislæsere[...],2018), which might force newspapers to reflect public opinion even more, it is particularly interesting to re-investigate these two newspapers. Not only in another topic, to see if other topics differ just as much in its framing. But also to see if a growing public interest influence one or both newspapers in comparable ways over time, with respect to number of articles posted and the sentiment of those articles.

To capture these questions, this paper seeks to investigate if there is a difference in the development of a key topic in the two newspapers 'Information' and 'Jyllandsposten'. The topic is the growing debate concerning diet, more specifically, the discussion of vegetarianism as an alternative to carnism.

All over the world it has become more and more normal to restrain ones meat consumption. To give an illustration of this, the amount of vegan people in the US has increased by 500 % over the last three years (Top trends[...], 2017). In denmark the trend shows as well; 'Dansk Vegetarforening' (the Danish vegetarian society) pronounced in 2016 that they had seen a

[1] This study is written unitedly by Sebastian and Nina. All sections and subsections have been touched upon by both minds and we have therefore chosen not to write individual names next to each and every title.
Moreover, footnotes will from this moment on serve as indicators gently suggesting the reader to look to the endnotes for further explanation in regard to the concepts and programs that are being dealt with.

65% increase in their membership base through the course of five years (Danskerne spiser mindre kød, 2016). Moreover there has been a decrease in the danish meat consumption in the time period from 2004 to 2011; where the average danish citizens would devour the ample amount of 116 kilograms of meat a year in 2004, the consumption was only 86 kilograms per capita in 2011(Økonomisk analyse, 2014). Furthermore this decrease in meat consumption is coupled with a growing demand for vegetarian products in the supermarkets as shown by the sales statistics of the wholesale distributor COOP. (Danskernes madvaner, 2006). What is more is that the danish dietary recommendation changed in 2013 to embrace a more plant-based diet and advise a more restrictive consumption of meat and dairy.(De officielle kostråd, 2018)

However Denmark is an agricultural country with an extensive production centered around livestock farming; having a pig population of 12,5 mio, double the amount of citizens living in all of Denmark (1. juli 2017)(Svinebestand[...], 2017). So the questions are: will the rising interest in vegetarianism show in the newspapers? Does the shift in public opinion reflect in the media's coverage of the topic of vegetarianism?

1.1 Media's influence on public opinion

As presented above vegetarianism is of growing interest to the public, and accordingly this paper is interested in the interplay between these public opinions and the media's different outlets. Studies show that newspapers carefully choose which statistics and facts to present in order to prime their readers in a wanted direction (Balahur & Steinberger 2009; Balahur et al. 2013; McCombs & Shaw, 1972). For instance as seen in the example with the Iracy war mentioned earlier in the introduction. Each and every newspaper frames the given topic to their readership. With framing in mind it is important to clarify that the media and its way of framing have a noticeable effect on individuals and thereby the society and its culture (Perse & Lambe, 2017). For instance, a positive correlation was found between the amount of cosmetic surgery makeover publications and the young adults in the US's desire to undergo cosmetic surgery (Nabi, 2009). Another example is the medias hinderince of the Feminist movement in the US. A study outlines the detrimental effects the media's anti-feminist frames had on the movement by affecting the public at large (Terkildsen & Schnell, 1997).

Nevertheless this media-framing can go wrong, as seen in a study concerning framing of animals rights in the 'Go veg' campaigns in the US. This study showed that the animal rights

activist must find a balance between; framing their movement as a logically better choice for the environment, which was found to be most convincing, and showing violent footage of animal cruelty, which has a very polarizing effect on some crowds (Freeman, 2010).

All things considered, with the rise of the internet and alternative sources of news next to newspapers, the newspapers can be expected to also mirror the public's opinion, more than being the more strictly agenda-setting machine. As the public are becoming more prone to select other sources of information as their choices have multiplied, it is interesting to see if the newspapers might be forced to mirror the public opinion. Due to this, this study investigates if this change in the danish public opinion, concerning vegetarians and carnivores, reflects in the danish medias. To investigate how these opinions express themselves, sentiment analysis is used as a tool.

1.2 Sentiment analysis

Sentiment analysis, also known as opinion mining, is used to deal with opinions, emotions and attitudes by looking at the sentiment and subjectivity of texts (Pang & Lee, 2008). As a tool it permits an analyst to investigate the emotional valence of individual words in larger texts or corpora. This allows the analyst to understand how individuals think and behave under the assumption that the public and the media mirror each other, as premeditated in the section above. It is a way for the analyst to glimpse into the inner workings of individuals cognition and thereby society's collected cognition.

As Natural Language Processing (NLP)[4] it is used in several different contexts such as computer science, social science and management science. In the business world sentiment analysis has become important, as it is the first tool in history that allows investigations into vast online datasets to find opinions through a computational treatment (Liu, 2012). Companies can develop new products or services in the light of what they find out about the general public opinions. The public/customers are also interested in knowing the opinions of others when shopping for new product or meditating which services to buy (Liu & Zhang, 2012).

Sentiment analysis is for instance used to predict the success of movies by looking at the positive sentiment on weblogs (Mishne & Glance, 2006). To understand what kind of framing a campaign like the 'Go Veg' should choose to optimize its influence(Freeman, 2010). Or to classify customers reviews of hotels (Gräbner, Zanker, Fliedl & Fuchs, 2012),

which allows new customers to discover trends in customers opinions and herby carefully consider which hotel to choose. Herewith it is no surprise that sentiment analysis is one of the fastest growing research areas in computer science, as shown in a study dealing with the evolution of sentiment analysis (Mäntylä, Graziotin & Kuutila, 2016).

Although it is easier to find the sentiment in texts of more extreme attitudes, like the campaign for 'Go Veg' or twitter feeds, sentiment analysis is in this study used to find a public opinion concerning vegetarianism and carnism in the aforementioned newspapers 'Information' and 'Jyllandsposten'.

1.3 The Danish dictionary AFINN

AFINN is the sentiment dictionary used in this study. It is created by Finn Aarup Nielsen, an Associate Professor at DTU Compute specializing in data mining, (Nielsen et al., 2011) and chosen for this study under the assumption that it is notably different from other sentiment dictionaries in being Danish, without being a Google Translated version of a foreign language dictionary. This avoids mistranslations and the implications that stem from cultural differences. For instance is the word "tax" rated as negative in the translatable canadian NRC emotion Lexicon (NRC emotion lexicon, 2018), while the AFINN gives it a positive rating.

The dictionary consists of 3552 individually scored words on a scale ranging from a negative (-5) to a positive (5), all labeled by the creator Finn Aarup Nielsen himself (See discussion for why it is a limitation only having one annotator). Not all of these words are unique, some are mere inflections of the same word as is the case with 'sult' (hunger) and 'sulte' (starve). Worth noting is also the fact that the AFINN dictionary has a greater quantity of negatively scored words. 65% of the glossary is negative. (Nielsen et al., 2011).

The-bag-of-word method is foundational to this dictionary meaning it disregards both grammar and word order. Instead it looks at a text as a collection of words awaiting individual scoring. With such a method in use, some preprocessing needs to be put in place (See Analysis). An obvious advantage of the AFINN dictionary is that it includes inflections of the same word, so that the preprocessing does not need to cut words down to their stem, e.g. cutting away 'er' (s) from 'elsker'(loves) to arrive at 'elsk' (love).

To give an illustration of how AFINN practise the bag-of-word method a passage from an article, in the corpus regarding the vegetarian keyword 'vegetar' in the newspaper Information, has been selected.

Jeg er en egovegetar. Personligt er jeg helt med på, at de fleste af mine holdninger er hykleriske(-3), når jeg stadig køber smør og elsker (2) mozzarella. For at være helt ærlig (2), så skyldes en betragtelig del af mine bevæggrunde for at være vegetar ren (2) egoisme(-3): Jeg har altid mistænkt (-1) kød - og især rødt eller forarbejdet kød - for at være relativt uforeneligt med den menneskelige krop.
(Information 2014-03-08)

I am an ego-vegetarian. Personally, I totally agree that most of my attitudes are hypocritical (-3) when I still buy butter and love (2) mozzarella. To be completely honest (2), it is so that a significant part of my reasons for being a vegetarian is sheer (2) selfishness (-3): I have always suspected (-1) meat - and especially red or processed meat - to be relatively incompatible with the human body.

As seen above, words with a positive sentiment are rated with a positive number from (1) to (5) and negative words are rated with an negative number from (-1) to (-5), while words that are not included in the AFINN dictionary get a score of 0. The MeanSentimentScore (M.S.S.) for this passage is (-0.015), which is a fairly weak negative M.S.S. considering the fact that the overall mean for the corpus used in this study range from a negative (-0.10) to positive (0.12). With this in mind it is also worth nothing is that the AFINN dictionary scores all words not encompassed in its library with a neutral 0, which explains why the M.S.S. is in decimal numbers. The passage above is just a snippet of a greater whole, but the M.S.S. ends up capturing the ambivalence of choosing the vegetarian lifestyle compared to a vegan one, as described by the author. The complete articles tend to show a stronger overall positive or negative M.S.S. However one single article does not have much to tell, but on the other hand comparing across a corpus of 800 articles it is possible to detect trends and find patterns in the different degrees of M.S.S.
Dictionary based sentiment studies, using a sentiment scales ranging from positive to negative, like AFINN, have showed impressive results reaching accuracy levels ranging from 59% to 75% across different datasets when put up against manual annotators (Koto & Adriani, 2015). It is almost as accurate as trained manual annotaters who only reached inter-

annotator agreement in 82% and 81% cases in two studies respectively using similar scales ranging from positive sentiment to negative (Wilson et al., 2005; Balahur et al., 2013). With this in mind it can be confidently assumed that whatever accuracies concerning sentiment, that will be found in this paper, will have an inherent accuracy level of under 80%, since sentiment analysis' will not be able to surpass the human level inter-annotator agreement it is based on.

1.4 Hypotheses

All things considered the three hypotheses of this paper are the following:

H1) There is a significant development in positive sentiment score for the vegetarian keyword 'vegetar' and a decline in sentiment score for the carnivores keywords 'kød/spiser' over a ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017.
H2) There is a significant difference in the development of sentiment for the vegetarian keyword 'vegetar' and carnivores keywords 'kød/spiser' in the two newspapers 'Information' and 'Jyllandsposten' over a ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017.
H3) As a representation of the growing interest in the vegetarian diet, the amount of articles posted by the newspapers 'Information' and 'Jyllands Posten' has increased over a ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017.

2. Materials and Methods

To investigate the hypotheses a corpus of 800 danish newspaper articles were assembled from the online database Infomedia[5] using web-mining. The articles were divided into four groups, namely split evenly between the two newspapers 'Information' and 'Jyllandsposten' and also evenly between the danish keywords: 'vegetar' (vegetarian) and 'kød'/'spiser' ('meat'/'eating') (See Table 1). The keywords were chosen in order to capture the most relevant articles in respect to the rise of vegetarianism; 'vegetar' were for instance used

instead of 'veganer' (vegan) as it was used to a far greater extend in both newspapers, which enabled the study to make better use of Infomedias relevance function. This function works by looking at the frequency of the chosen keywords across the whole database amongst other scoring mechanisms[6].

Additionally the word 'spiser' (eating) was added to the word 'kød' (meat) in order to capture the articles dealing with meat in the right context. 'Kød' on its own occurred in various peculiar ways; for instance in the proverb: "i kød og blod" (in flesh and blood). After choosing the keywords, 20 articles were collected for increments of one year in the period from 07.12.2007 to 07.12.2017 in each of the four groups. This procedure for compiling the corpus was done in order to make developments over time worthy of comparison across all four groups.

In the search for articles, articles containing less than 125 words were excluded. These articles were excluded on the basis that they generally tended to be mere references of other articles and thereby not relevant to the sentiment study. Moreover, because the articles were so short, their M.S.S. could more easily be influenced by the positive or negative scoring words (See section 'The Danish dictionary AFINN'), making the corpus prone to outliers.

Keywords / Newspaper	Information	Jyllands Posten	Total number of articles
'vegetar'	200	200	400
'kød' & 'spiser'	200	200	400
			800

Tabel 1: Division of corpus between newspapers and keywords

3. Analysis

In order to test the hypotheses on the corpus, the articles were scored using the AFINN sentiment dictionary. That is to say; 20 articles from each of the four groups were downloaded for each increment of one year in the period from 07.12.2007 to 07.12.2017. The articles were pulled from Infomedia in PDF-format, yet quickly transformed into txt-files using the online tool Pdftotext[7] for easy processing. Firstly each txt-file, containing 20 articles, was run through a python script using Spyder 3[8] that artificially split the text into

separate articles. This was done under the assumption that articles inclined to split naturally around every 50th line and with the intention of automating the process. Although this method resulted in anything from 20 to 49 articles, this did not affect the overall analysis because the study looks at the M.S.S. for each year in total. Secondly the script cleaned the newly split articles; removing all punctuation and all copyright infringement messages, and thirdly converting all characters to a lowercase - the standard used by the AFINN dictionary. Finally all characters were tokenized for individual scoring. All individual words were then assessed and scored up against the AFINN dictionary before the python script returned reassembled articles in a CSV-file format.

With the scoring in mind, a M.S.S. was calculated as opposed to an absolute sentiment score. This was done in order to make the articles comparable across all four groups, although they might differ in total sum of scored words. Moving on, the hypotheses were tested on the newly scored corpus in the RStudio environment[9]. Using the 'M.S.S.' as the dependent variable for the first and second hypotheses and 'frequency' for the third. The three hypotheses were tested with linear regression with both 'time', 'newspaper' and 'keywords' as independent predictors. All models conformed to assumptions as tested by residual plots.

Additionally the second hypothesis H2 was followed up with an explorative visualisation of the data to examine why the the dependent variable would not succumb to a linear model.

4. Results

The corpus did not conform to the first hypothesis H1. The predictors: The keywords ('vegetar' and 'kød/spiser') and time are not significant predictors for the M.S.S.: $\beta = -0.0007$ (SE =0.0005), $t = -1.54$, $p = 0.13$, in a two-way interaction linear regression model, since $p > 0.5$. Nevertheless a scatterplot with a fitted linear regression line (Figure 1) demonstrates the relationship between the linear development in sentiment over time for both keywords ('vegetar' and 'kød/spiser') respectively.

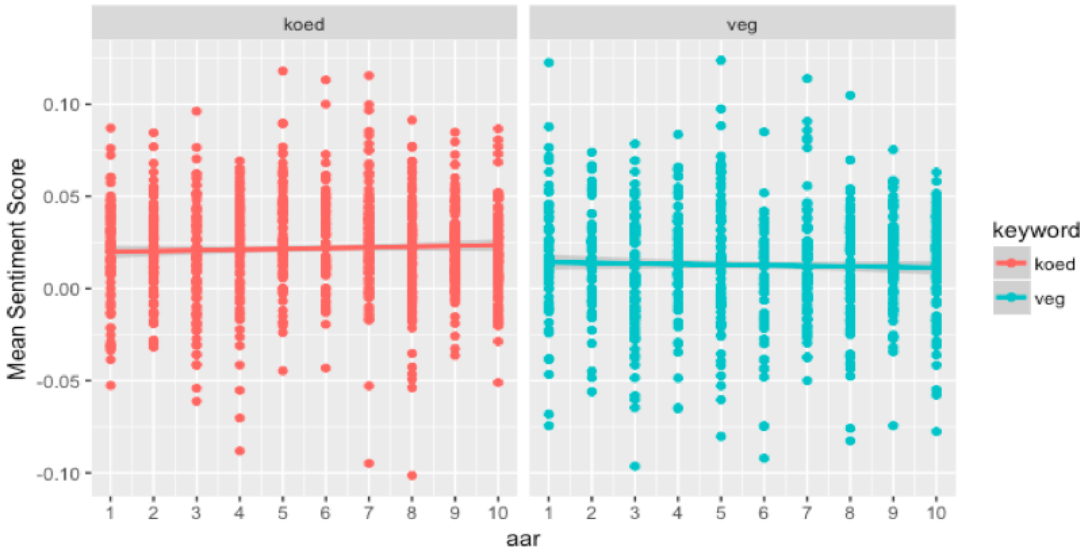

Figure 1. Keywords M.S.S. over a ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017.

Similarly the corpus did not conform to the second hypothesis H2. Time, keywords and the selected newspapers were not significant predictors for the sentiment: $\beta = 0.0003$(SE =0.001), t =0.33, p = 0.74, in a three-way interaction linear regression model, since $p > 0.05$. In other words, the relationship found between the framing of newspapers for the keywords over a ten year period, is insignificant. Yet a scatter plot with a fitted linear regression line (Figure 2.) demonstrates the relationship between the linear development in sentiment over time for both keywords respectively.

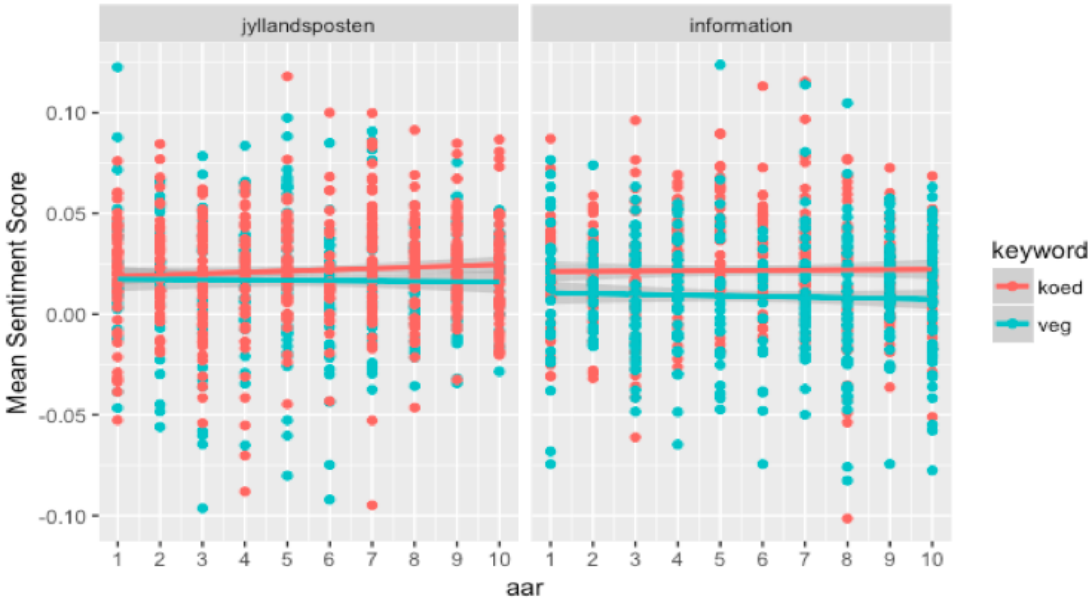

Figure 2. Keywords M.S.S. for newspapers over a ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017.

Finally the Corpus did not confirm to the third hypothesis H3, dealing with the frequency/number of articles for each keyword across all national Danish newspapers over the ten year period. There is no significant effect when looking at the keyword 'vegetar': $\beta = 5.19$ (SE = 5.51), $t = 0.94$, $p = 0.37$, in a linear regression model. Likewise there is no significant effect when looking at the keywords 'kød' and 'spiser': $\beta = 5.61$ (SE = 4.47), $t = 1.26$, $p = 0.25$ in a linear regression model. A scatter plot with a fitted linear regression line (Figure 3) demonstrates linear development in frequency over time.

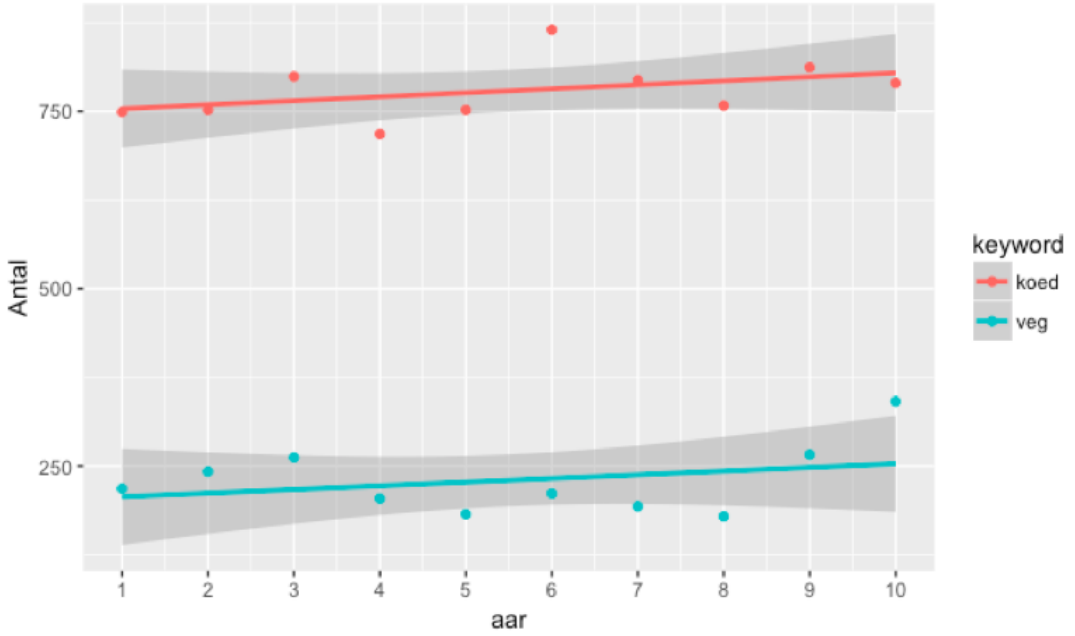

Figure 3. Frequency of articles concerning keywords over a ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017.

4.1 Post Hoc

In the light of the fact that none of the hypotheses were found to show anything significant, a follow-up explorative visualisation of the corpus was created. Looking at (Figure 4.) it becomes clear why it is hard to fit a linear regression model to the sentiment of the two newspapers. To give an example; the sentiment for the carnivorous keywords peaked in the period 2011-2012, (fifth year in (Figure 4) in Information) and starts descending thereafter, hinting at the need for non-linear regression and a curved slope.

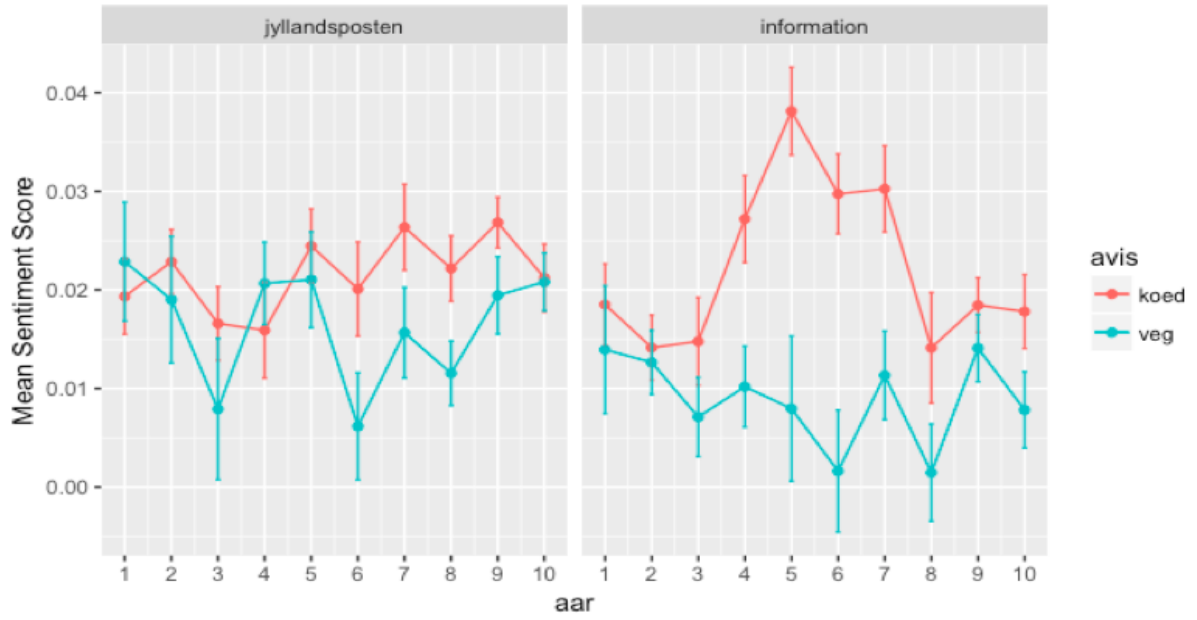

Figure 4. M.S.S. with errorbars for all four groups over a ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017.

5. Discussion

The results imply that there are no significant linear developments in M.S.S. over time across the Danish newspapers 'Information' and 'Jyllandsposten' for the keywords 'vegetar' and 'kød/spiser'; therefore neither H1 or H2 were confirmed. Moreover H3 was not confirmed either, signifying that there is no significant linear development in the frequency of national Danish newspaper coverage of the keywords over time.

Nonetheless an explorative follow-up visualisation of the corpus reveal why the study fails to get significant results with linear models. The data is simply not linear in nature. What the explorative visualisation can also say is that there seems to be a general positivity bias for both newspapers towards these food-related topics. With an M.S.S. range from -0.10 to 0.12, the overall mean is 0.02. Which is noteworthy since the AFINN dictionary contains a greater amount of negatively scored words (65%) (Nielsen et al., 2011). Furthermore it displays a notably more positive coverage of carnivorous keywords as seen in (Figure 4), year 2012-2016 in Jyllandsposten (Year 6-9) and 2010-2015 in Information (Year 4-8) . Looking at the overall M.S.S. for the carnivorous keywords and vegetarian keyword respectively, it is

apparent as well, with a M.S.S. of 0.02 for the carnivorous and a M.S.S. of 0.01 for the vegetarian.

Altogether this study did not find the linear patterns it sought to find, but though it might have been due to the fact that there was no linearity to begin with, there is a myriad of methodological limitations that undoubtedly messed with the analysis and that could be taken into account in future research.

5.1 Methodological Limitations

5.1.1. Sentiment analysis and the bag-of-words method

Using sentiment analysis and especially the bag-of-words method hold obvious disadvantages of not being able to pick up on syntax such as negations: "Det er **ikke godt** for debatten om den **manglende dyrevelfærd**" (It is **no good** for the debate concerning **the lack of animal welfare**) (Information, 2011-12-23) . Likewise irony and sarcasm are neither accounted for by sentiment dictionaries like AFINN, that might seem otherwise obvious to human annotators. In one study the annotators reached an inter-annotator agreement of 80% for what constitutes sarcasm based on contextual cues (Tepperman, Traum & Narayanan, 2006); close to the threshold for human annotator agreement of around 80% as brought up in the introduction. Moreover the bag-of-words method does not account for semantics either. That is to say that an article can be positively biased toward a given subject although it is using negative language. For instance; if an article favorably disposed towards vegetarianism expresses thoughts about the effects of meat-consumption on climate change, as seen in this example: "Desuden lider vores miljø og klima stærkt under produktionen af animalske fødevarer, som er og bliver fuldstændig ubæredygtig for både miljø og mennesker." (Moreover, our environment and climate suffer greatly from the production of animal food products, which is and stays completely unsustainable for both the environment and humans.) (Information, 2017-11-30). The bag-of-words method can only interpret this as a general negative tone in the article, leading to a lower overall M.S.S. for the vegetarian keywords, because the automatic sentiment analysis cannot see the structure nor the overall meaning. As an additional example of categorization mistakes, two of the most relevant articles, as rated by Infomedia, titled "Helt uden kød!" (Wholy without meat!) (Information, 2012-03-03) and

"Mindre kød - flere bælgfrugter!" (Less meat - more legumes!) (Information, 2012-18-02) are presented. These articles are pro-vegetarian, but got stuck in the carnivorous keyword group, which reveal how many of the vegetarian articles highlighting the positive sides of the vegetarian diet were not showing up in the vegetarian keyword group. Which in part also assist in explaining why there is a greater positivity bias of meat in Information 2011-2012 (Year 5, Fig. 4) as touched upon earlier. In line with this the keywords 'vegetar' and 'kød/spiser' are sources of error. They do not narrow down the categories enough. Other words like 'vegetarisme, veganer, veganisme'(vegetarianism, vegan, veganism) etc. could have been chosen for the vegetarian category to help Infomedia filter and rank the most relevant articles. Additionally words could have been excluded, as for instance 'kød' in the vegetarian category, though this could cause problems in our case, as most articles dealing with this category mention the word as a part of the debate. If all articles were divided correctly into categories it might have been possible to show this aforementioned trend and change in public opinion of becoming vegetarians.

5.1.2. AFINN

As stated in the introduction, sentiment analysis is often used for large corpus of data such as twitter feeds. AFINN was for the most part created for this very purpose, namely microblogging (AFINN, 2018). This is a limitation. A substantial amount of the AFINN dictionary are words often used in tweets, but seldom in actual newspaper articles. Tweets are of another caliber than articles; they contain a maximum of a 140 characters and use words like 'blah', 'blockbuster' and 'boostet', which are all included in the Danish AFINN dictionary. However, what is not included are words such as 'bæredygtig'(sustainable), 'undertrykkelse'(suppression) and 'holdbare'(durable), which are likely to show up in newspapers that do not limit the quantity of characters nor use language/slang in the same way. This also shows in the example from the introduction, section "The Danish dictionary AFINN", where the word "uforenligt" (incompatible) is not rated although it is a strongly negative word. Furthermore the dictionary, as stated in the introduction, does not hold 3552 unique words, but a huge amount of inflections. To give an illustration, there are six inflections of the word 'fattig' (poor), again indicating a limitation of the scope of the dictionary.

With the limitations of the dictionary in mind, it is also important to note that the whole dictionary has been rated solely by Finn Aarup Nielsen, as premeditated in the introduction. Since no additional annotators has been reviewing, any chance of inter-annotator agreement and its inherent validity cease to be possible. This is a huge source of error in the AFINN dictionary, as studies described in the introduction showed that trained manual annotaters only reach inter-annotator agreement in 82% and 81% respectively using a similar scale ranging from positive sentiment to negative (Wilson et al., 2005; Balahur et al., 2013).

5.2. Further research

5.2.1. An improvement to AFINN

Steps that could be taken to remedy the outlined flaws of the AFINN dictionary include increasing its scope to include more words apparent in newspaper articles. This would prove a remarkably better dictionary, but not fix the complications of the bag-of-words approach. Here a Context-based algorithm, which has proven an improvement in other studies(Sharma, Chakraverty, Sharma & Kaur, 2017), might help with the scoring; for instance making negations shift the sentiment of the following word from positive to negative and vice versa, or by making adverbs, such as ‘highly’ and ‘really’, adapt the sentiment in similar fashion. Additionally, one could treat all quotes and citations in a separate manner to avoid the issue of syntax, where negative citations or highly emotional references is not matching the overall tone of on article. Not unlike the example given earlier, where an article favored vegetarianism yet expressed negative thoughts on the effects of meat-consumption. Another important source of error that could be worked on in future research, is the complications regarding sarcasm. A potential way of dealing with this could be to use Machine Learning and train an algorithm on an extensive dataset to recognize these linguistic maneuvers, as has shown effective at predicting sarcasm by 77% - a significant improvement from chance level (Davidov, Tsur, & Rappoport, 2010). Alternatively the filtration process (See Analysis) could also be expanded and set-up to exclude all obvious sarcastic remarks, by removing all words and sentences marked with quotation marks - a universal clue for sarcasm.

5.2.2. The dictionary NRC

As premeditated in 'The Danish dictionary AFFIN' section, other dictionaries could contribute with new important factors to the analysis. Another dictionary that could be used is the canadian dictionary NRC, which contain 4671 danish words - 1119 more than the AFINN dictionary. As well as including more words, the NRC rates all words in 6 different emotions: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust simultaneously with rating them on a range from positive to negative. This could be an advantage when for instance investigating more directly how the different newspapers choose to frame the given topics. Instead of only looking at framing in a spectrum going from positive to negative sentiment, NRC would enable a study more narrowly focused on the emotions in regard to the public's opinions. If for instance taking the aforementioned "Go Veg" campaigns, this dictionary would enable the study to look at e.g. framings with anger and fear versus framing with joy and trust. Yet, as classified in the introduction section 'The Danish dictionary AFINN' the NRC is limited by the fact that it uses Google Translate and therefore might be prone to errors that stem from cultural differences of interpretation of words. It might therefore contain mistranslated words and it can misread the context of the specific words. To correct this limitation several danes could go through all of the 4671 words in the NRC dictionary and review their translations and rating together to arrive at a inter-annotator agreement. This would enable the NRC dictionary to take cultural differences and contexts into consideration, like in the aforementioned example with "tax" seen in the introduction section The Danish dictionary AFINN (NRC emotional lexicon, 2018).

5.2.3. Better categories

As mentioned earlier, a great deal of the 800 articles used in this study were put in the wrong category, thus an article about vegetarianism like the aforementioned "Mindre kød - flere bælgfrugter!" (Less meat - more legumes!) (Information, 2012-18-02) could end up in the carnism category. In future research topic modelling could be used as another text-mining tool. This would enable the investigation to divide the articles more accurately into the correct categories, as each article would have its most frequent words shown. Additionally, this method would pick up on articles dealing with irrelevant topics, like recipes.

Another way to overcome the problems regarding categorizing could be to manually choose each an every article from Infomedia into the right category. This would be an advantage since the manual web-miners would be able to pick up on both sentiment and sarcasm. All things considered it would be interesting to look at the distinction between a computational treatment of the web-mining, using Infomedias relevance function, and a manual treatment.

5.2.4. Expanding the study

The results shown in (Figure 3) are the general frequencies for the two keywords in all national Danish newspapers, but to expand this study it could be interesting to look at frequency more specifically for both 'Information' and 'Jyllandsposten' used in this study. For instance, assuming that the left-winged newspaper 'Information' has a more positive attitude toward vegetarianism, which should be stressed - could not confirmed in this study, it could be investigated whether the newspaper published more articles about vegetarianism; and vice versa for Jyllandsposten and carnivorous articles.

To expand the scope of the study it could also be interesting to look at the frequency of posts/articles or amount of events/groups in other online databases, such as twitter and facebook. To compare any linear development found here up against the frequencies found for the chosen newspapers. This comparison would be a direct way to see if public opinion and the medium of newspapers mirrored each other on this topic, as touched on in the introduction.

In regard to this, the different newspapers readerships could also be examined; a danish study shows that the online media has a much younger readership than newspapers (Schrøder, 2015). Furthermore juveniles are more prone to become vegetarians or vegans(Vegan profile, 2018). It would be interesting to investigate whether or not this played a role in relation to why this study did not find any significant linear developments. Could vegetarianism simply be a phenomena of growing interest outside the chosen newspapers readerships?

As mentioned in the introduction, section Sentiment Analysis, the online media contains a huge amount of data revealing public opinions, e.g. movie and restaurant reviews. These datasets could also contribute to an expansion of this study; Do the Danes rate vegetarian restaurants better today than ten years ago? And what is the sentiment of comments to movies framing vegetarianism in either a positive or a negative light?

6. Conclusion

This study analysed a corpus of 800 articles split equally into four groups, separating articles by the two newspapers 'Information' and 'Jyllandsposten' and by articles concerning either the vegetarian keyword 'vegetar' or the carnivorous keywords 'spiser' and 'kød'. In an attempt to showcase a linear development in each group across the ten year period ranging from the 7th. of December 2007 to the 7th. of December 2017, the sentiment dictionary AFINN was used to collect a mean sentiment score for each article to construct linear regression models. The models showed no compelling evidence for the predicted linear development as no significant effects were found in any of the models. There were no significant decline in sentiment score for the carnivorous keywords over time, neither were there any significant positive development in sentiment of the vegetarian keywords, nor when zooming in on the individual newspapers. Additionally the study attempted to show a positive linear development in the amount of articles posted in national newspapers across the ten year period. Nonetheless no significant development was found.

The main contribution of this study is to use sentiment analysis to get a more impartial and quantitative measure of the development in public opinion concerning vegetarianism as reflected in the sentiment of the written medium, newspapers - an unsuccessful endeavour into seeing whether traditional media catch on to topics like vegetarianism that are of growing interest to the public. Noticeable limitations show up when inspecting the bag-of-words approach and the AFINN dictionary and this studies procedure for web-mining, which call for future research with expectantly more success. Future studies should mainly focus on getting syntax and semantics right, along with possibly expanding the grasp of the study by including more data, e.g. in the form of facebook posts and restaurant reviews.

References

Davidov, D., Tsur, O., & Rappoport, A. (2010). Semi-supervised recognition of sarcastic sentences in twitter and amazon. Paper presented at the *Proceedings of the Fourteenth Conference on Computational Natural Language Learning,* pp. 107-116.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51-58.

Freeman, C. P. (2010). Framing animal rights in the "Go veg" campaigns of US animal rights organizations. *Society & Animals, 18*(2), 163-182.

Gräbner, D., Zanker, M., Fliedl, G., & Fuchs, M. (2012). *Classification of customer reviews based on sentiment analysis* na.

Hjarvard, S. (2007). Den politiske presse-en analyse af danske avisers politiske orientering. *Journalistica-Tidsskrift for Forskning i Journalistik, 2*(5)

Koto, F., & Adriani, M. (2015). A comparative study on twitter sentiment analysis: Which features are good? Paper presented at the *International Conference on Applications of Natural Language to Information Systems,* pp. 453-457.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. *Mining text data* (pp. 415-463) Springer.

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2016). The evolution of sentiment analysis-A review of research topics, venues, and top cited papers. *ArXiv Preprint arXiv:1612.01556,*

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly, 36*(2), 176-187.

Mishne, G., & Glance, N. S. (2006). Predicting movie sales from blogger sentiment. Paper presented at the *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs,* pp. 155-158.

Nabi, R. L. (2009). Cosmetic surgery makeover programs and intentions to undergo cosmetic enhancements: A consideration of three models of media effects. *Human Communication Research, 35*(1), 1-27.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *ArXiv Preprint arXiv:1103.2903,*

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1–2), 1-135.

Perse, E. M., & Lambe, J. (2016). *Media effects and society* Routledge.

Schrøder, K. C. (2015). *Danskernes brug af nyhedsmedier 2015: Brugsmønstre og nyhedsrepertoirer 2015* Roskilde Universitet.

Sharma, S., Chakraverty, S., Sharma, A., & Kaur, J. (2017). A context-based algorithm for sentiment analysis. *International Journal of Computational Vision and Robotics, 7*(5), 558-573.

Tepperman, J., Traum, D., & Narayanan, S. (2006). " Yeah right": Sarcasm recognition for spoken dialogue systems. Paper presented at the *Ninth International Conference on Spoken Language Processing,*

Terkildsen, N., & Schnell, F. (1997). How media frames move public opinion: An analysis of the women's movement. *Political Research Quarterly, 50*(4), 879-900.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. Paper presented at the *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing,* pp. 347-354.

Xiao, W. H. (2011). On the network emergent sentiment and its supervision. (Master, Hunan Normal University (People's Republic of China)). *PQDT - Global,* Retrieved from https://search-proquest-com.ez.statsbiblioteket.dk:12048/docview/1874854232?accountid=14468

Web pages

AFINN - brede wiki Retrieved 1/4/2018, 2018, from http://neuro.imm.dtu.dk/wiki/AFINN

Danskernes madvaner 2016_Coop analyse.pdf Retrieved 1/4/2018, 2018, from https://om.coop.dk/Upload/om.coop.dk/Publikationer/analyser/Danskernes Madvaner 2016_Coop Analyse.pdf

NRC emotion lexicon Retrieved 1/4/2018, 2018, from http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

De officielle kostråd - alt om kost Retrieved 1/4/2018, 2018, from http://altomkost.dk/deofficielleanbefalingertilensundlivsstil/de-officielle-kostraad/

Økonomisk analyse - danskernes-forbrug-af-koed-2014-2.pdf Retrieved 1/4/2018, 2018, from http://www.lf.dk/tal-og-analyser/forbrugere-og-trends/forbrug-og-detail/markedsanalyse-om-koed-2014

PDF to text – convert PDF to text online Retrieved 1/4/2018, 2018, from http://pdftotext.com/

Vegan profile Retrieved 1/4/2018, 2018, from http://www.imaner.net/panel/profile.htm

Journalistic objectivity. (n.d.). Retrieved 1/4/2018, 2017, from https://en.wikipedia.org/wiki/Journalistic_objectivity

Færre avislæsere og færre læste aviser Retrieved 1/4/2018, 2018, from http://www.dst.dk/da/Statistik/nyt/NytHtml?cid=24083

Top Trends in Prepared Foods 2017: Exploring trends in meat, fish and seafood; pasta, noodles and rice; prepared meals; savory deli food; soup; and meat substitutes. Retrieved 1/4/2018, 2018, from https://www.reportbuyer.com/product/4959853/top-trends-in-prepared-foods-2017-exploring-trends-in-meat-fish-and-seafood-pasta-noodles-and-rice-prepared-meals-savory-deli-food-soup-and-meat-substitutes.html#free-sample

Danskerne spiser mindre kød Retrieved 1/4/2018, 2016, from https://www.b.dk/nationalt/danskerne-spiser-mindre-koed

Svinebestand ligger stabilt omkring 12,5 mio. svin Retrieved 1/4/2018, 2016, from http://www.dst.dk/da/Statistik/nyt/NytHtml?cid=23460

Endnotes

1. *Carnism* - "Carnism is a concept used in discussions of humanity's relation to other animals, defined as a prevailing ideology in which people support the use and consumption of animal products, especially meat."
Retrieved 1/4/2018, 2016, from
https://en.wikipedia.org/wiki/Carnism

2. *Vegetarianism* - "Vegetarianism is the practice of abstaining from the consumption of meat , and may also include abstention from by-products of animal slaughter."
Retrieved 1/4/2018, 2016, from
https://en.wikipedia.org/wiki/Vegetarianism

3. *Framing* - "To frame is to select some aspects of perceived reality and make them more visible in a communicating text to promote a particular definition of a problem, the causes of the problem , the moral evaluations and / or solutions to the described phenomenon."
Retrieved 1/4/2018, 2016, from
https://da.wikipedia.org/wiki/Framing

4. *NLP* - "Natural language processing (NLP) is a field of computer science [...] concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to fruitfully process large natural language data."
Retrieved 1/4/2018, 2016, from
https://en.wikipedia.org/wiki/Natural_language_processing

5. *Infomedia* - "A Danish media surveillance company that collects and monitors content from all national and regional newspapers."
Retrieved 1/4/2018, 2016, from
https://infomedia.dk/mediearkivet/

6. *Infomedias relevance function* - "Scores by number of occurrences of keywords in the article, keyword frequency in comparison to other words in the database, the number of keywords in the same article and length of article."
Retrieved 1/4/2018, 2016, from
https://infomedia.force.com/s/article/Hvad-er-relevansscoren

7. *pdftotext* - "Online tool for bulk conversion of PDF documents to plain text files."
Retrieved 1/4/2018, 2017, from
http://pdftotext.com

8. *Spyder* - "Spyder is a Scientific Python Development Environment"
Retrieved 1/4/2018, 2016, from
https://spyder-ide.github.io

9. *Rstudio* - "RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics."
Retrieved 1/4/2018, 2016, from
https://www.rstudio.com